A note on the statistical power in extended twin designs

Posthuma, D.; Boomsma, D.I.

published in
Behavior Genetics
2000

DOI (link to publisher)
10.1023/A:1001959306025

document version
Publisher's PDF, also known as Version of record

Behavior Genetics, Vol. 30, No. 2, 2000

A Note on the Statistical Power in Extended Twin Designs

Daniëlle Posthuma[1,2] and Dorret I. Boomsma[1]

Received 7 Oct. 1999—Final 20 Feb. 2000

The power to detect sources of genetic and environmental variance varies with sample size, study design, effect size and the statistical significance level chosen. We explored whether the power of the classical twin study may be increased by adding non-twin siblings to the classical twin design. Sample sizes to detect genetic and shared environmental variation were compared for kinships with only twins, kinships consisting of twins and one additional sibling, and kinships with twins and two additional siblings. The effect of adding siblings to the classical twin design was considered for univariate and bivariate analyses.

For the univariate case, adding one non-twin sibling resulted in a decrease in sample size needed to detect additive genetic influences in the presence of environmental influences. However, adding two additional siblings did not decrease the number of subjects as compared to the classical twin design. The sample size required to detect common environmental factors was also greatly decreased by adding one non-twin sibling. Adding two non-twin siblings resulted in a small additional decrease. In models including additive genetic, dominant genetic, and unique environmental effects, adding one sibling to a twin family decreased the required sample size to detect dominant genetic influences. Adding two siblings to a twin family resulted in only a slight additional decrease in sample size.

In the bivariate case a similar pattern of results was found, in addition to the observation that the overall required sample size, as expected, was lower than in the univariate case. The decrease in sample size from bivariate testing was more pronounced in a design with one or two additional siblings, as compared to a design with twins only. It is concluded that a well considered choice of family design, i.e. including families with twins and one or two additional siblings increases the statistical power to detect sources of variance due to additive and non-additive genetic influences, and common environment.

KEY WORDS: Sample size; heritability; methodology; sibship size; twin study.

INTRODUCTION

Recent advances in molecular genetics have made it possible to partition genetic variance into sources due to particular genetic loci (quantitative trait loci's; QTL's) and sources due to background genetic variance (Fulker, Cherny, & Cardon, 1995; Fulker, Cherny, Sham, *et al.*,1999; Nance and Neale, 1989; Boomsma and Dolan, 1998). A necessary first step in mapping complex traits to QTL's is to establish the amount of genetic variation that underlies the phenotypic variation of the trait. If phenotypic variation in a trait is found to be caused in part by genetic sources, linkage and/or association studies can be conducted in order to characterize the effects of specific genetic loci on the phenotypic variation. If phenotypic variation is not found to be heritable, the search for effects of specific genetic loci will not be initiated. However, in some cases it may be concluded that phenotypic variance in

[1] Department of Biological Psychology, Vrije Universiteit Amsterdam, Netherlands.

[2] Correspondence and requests for reprints should be sent to: D. Posthuma, Vrije Universiteit, Department of Biological Psychology, De Boelelaan 1111, 1081 HV, Amsterdam, The Netherlands. Telephone: +31 20 444 8814, telefax: +31 20 444 8832, email: danielle@psy.vu.nl

We gratefully acknowledge the financial support of the USF (grant number 96/22) and the HFSP (grant number rg0154/1998-B). We wish to thank Eco de Geus, Leo Beem and Conor Dolan for their comments on draft versions of this paper.

0001-8244/00/0300-0147$18.00/0

a trait can not be ascribed to genes because the statistical power to detect sources of genetic variation is insufficient (Svikis, Velz & Pickens, 1994; Pickens, Svikis, McGue, Lykken, *et al.*, 1991). This will preclude further searching for effects of QTL's on that particular trait, even though such QTL's may be present.

The statistical power of quantitative genetic studies is influenced by the size of the effect (e.g. heritability), the sample size, the probability level (α) chosen, and the homogeneity of the sample (Neale and Cardon, 1992; Cohen, 1992; Tanaka, 1987). Increasing the sample size is the most common way to increase the statistical power of a study, but is often limited by resources of time and money. Another means to increase statistical power is the use of multivariate testing. In the context of structural equation modeling the statistical power to detect genetic effects rises as a (non-linear) function of multivariate testing under the condition that the measures are correlated (Schmitz, Cherny, and Fulker, 1998). In the context of partitioned twin analyses it has been shown that choosing a different (e.g. other than 1 to 1) MZ to DZ ratio influences statistical power such that an MZ to DZ ratio of 1 to 4 is optimal for partitioned twin analyses (Nance & Neale, 1989).

In the present paper we focus on increasing the statistical power of the classical twin study by adding non-twin siblings to MZ and DZ twin pairs. Since non-twin siblings share on average half of their segregating genes, just like DZ twins, adding non-twin siblings to the classical twin design may provide an efficient way to increase the power to detect sources of genetic and shared environmental variance. Adding two more siblings to a twin kinship provides five additional observed covariances, whereas adding a whole new family consisting of two siblings provides only one additional observed covariance. In the present paper we examine the effects of adding non-twin siblings to twin families on the estimated sample size needed to detect additive genetic (A) variance (V_a), dominant genetic (D) variance (V_d), and common environmental (C) variance (V_c), with a power of 80% in the context of structural equation modeling.

METHOD

We calculated covariance matrices for three experimental designs, which differed in family constitution. Design 1 included only MZ twins and DZ twins. Design 2 included families with MZ and DZ twins and one additional sibling. Design 3 included families with MZ and DZ twins and two additional siblings. For all three designs we calculated the sample size needed to detect an effect of interest with a power of 80%. The MZ twins to DZ twins ratio was 1 to 1 for all three designs (thus, the ratio MZ to 'non MZ sibpairs', is not 1 to 1 for all designs). It should be noted that we report sample size in subjects and not in twin pairs. The same number of subjects refers to different numbers of twin pairs and a different number of families for all three designs. We will use the terms 'highest power' and 'fewest subjects needed' to refer to an optimal design to detect sources of phenotypic variance.

All analyses were carried out using the statistical software package Mx (Neale, 1997). Estimation of parameters was obtained by normal theory maximum likelihood. Goodness of fit testing was based on the likelihood ratio tests. First univariate models were considered. In order to obtain the sample size needed to detect varying levels of additive genetic variance with a fixed power level of $(1 - \beta) = .80$, covariance matrices were calculated with sources of additive genetic variance (V_a) accounting for 10% to 90% of the phenotypic variance in the presence of sources of common environmental variance (V_c) accounting for 00%, 10%, and 20% of the variance. Remaining variance was attributed to unique environmental (E) sources of variance (V_e). To detect sources of V_c covariance matrices were calculated with V_c accounting for 10% to 90% of the phenotypic variance in the context of sources of V_a accounting for 00%, 10%, and 20% of the phenotypic variance. In addition, covariance matrices were calculated with sources of variation due to A, D (dominant genetic variance) and E. Only the situation in which dominance was 'complete' (V_a to V_d = 2 to 1; see appendix I) was considered. In the ADE-models the total genetic variance, i.e. V_a and V_d together accounted for 30% to 90% of the total phenotypic variance. For all situations, remaining variance was attributed to V_e.

Since non-twin siblings, like DZ twins, share on average half of their genes, expectations for non-twin sibling covariances were modeled similarly to expectations for DZ covariances.

In the ACE-models the expected phenotypic variance (σ^2) of twins and siblings is $V_a + V_c + V_e$, the expected MZ covariance $V_a + V_c$, and the expected DZ and sibling covariance $.5\ V_a + V_c$. In ADE-models, the expected phenotypic variance is $V_a + V_d + V_e$, the expected MZ covariance $V_a + V_d$, and the expected DZ and sibling covariance $.5\ V_a + .25\ V_d$.

It is known that the use of a multivariate phenotype, as opposed to a univariate phenotype, results in a gain of statistical power if the multivariate traits are correlated (Schmitz *et al.* 1998). To find out how much

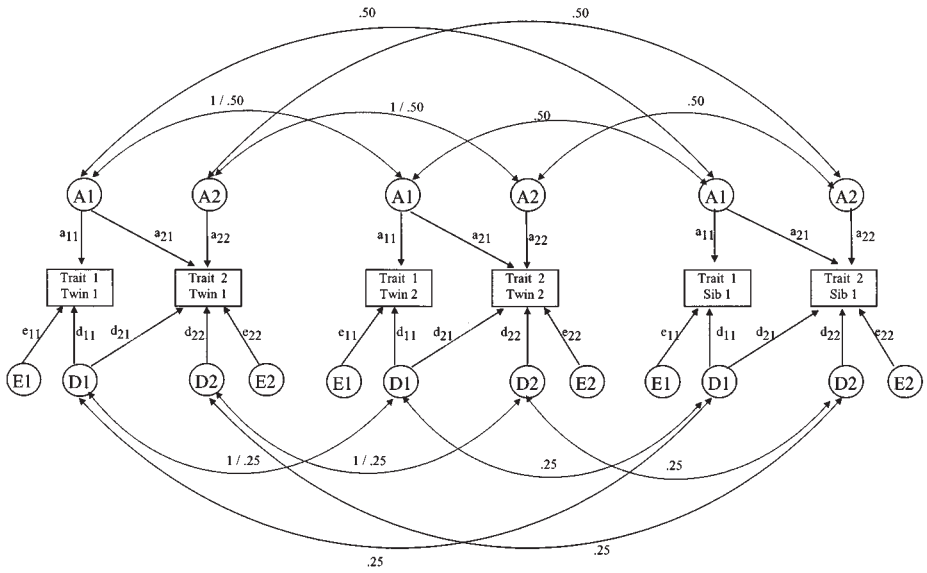


Fig. 1. Pathdiagram for the bivariate ADE-model, cholesky decomposition. Example for twins and one additional sibling, no unique environmental correlation (rE). The covariance between trait 1 and trait 2 is $(a_{11}*a_{21}) + (d_{11}*d_{21})$ and the correlation between trait 1 and trait 2 is $(a_{11}*a_{21}) + (d_{11}*d_{21})/\sqrt{(\sigma^2_1*\sigma^2_2)}$.

adding siblings *and* using a multivariate phenotype affects statistical power we also looked at several bivariate designs. We calculated covariance matrices for two traits with a phenotypic correlation of .50. Both traits could be influenced by A, C, and E or by A, D, and E. Total influences of sources of A, C or D, and E were uniform for each trait. The phenotypic correlation between the two traits could be due to additive genetic correlation (rA), dominant genetic correlation (rD), common environmental correlation (rC), or to unique environmental correlation (rE), depending on the specific situation that was considered. Figure 1 depicts the construction of covariance matrices for kinships consisting of twins and one additional sibling for a bivariate ADE-model (Cholesky decomposition) in which rE is absent and all phenotypic correlation is due to rA and rD. All latent variables have unit variance.

Power calculations were carried out by fitting the known model to the exact (population) covariance matrices as described in Neale and Cardon (1992). In models which contain a parameter which is known to be zero, the zero parameter can either be fixed at zero or freed (estimated) while computing the power to detect one of the other non-zero parameters. For example, when treating the ACE-model in which V_c is zero as an AE-model, the power to detect sources of variation due to A is significantly higher than when the ACE-model is treated as an ACE-model, i.e. with V_c estimated as a free parameter. In the power calculations the zero-parameter was estimated as a free parameter because we are interested in computing the power to detect V_a, in ACE-models, regardless of the value of V_c (and vice versa). The same reasoning applies to the bivariate calculations.

Constraining a certain set of parameters to zero and refitting the model provides the non-centrality parameter. From this non-centrality parameter the sample size required to reject the false model with a power of 80% and a significance level α of .05 can be calculated (Martin *et al.*, 1978; Hewitt and Heath, 1988) and is conveniently supplied by Mx.

RESULTS

Univariate Models

ACE-models

We fitted full univariate models with sources of variation due to additive genetic (A), common environmental (C) and unique environmental influences (E). Dropping either genetic or common environmental parameters and refitting the model provides the non-centrality parameter. With Mx (Neale, 1997) the corresponding number of subjects required to detect the parameter that was dropped with a power of 80% and α of 5% was calculated for 1 degree of freedom. Results concerning the estimated sample size (in subjects) needed to detect V_a in ACE-models for the three designs are depicted in Figure 2 (and appendix II). Figure 2a con-

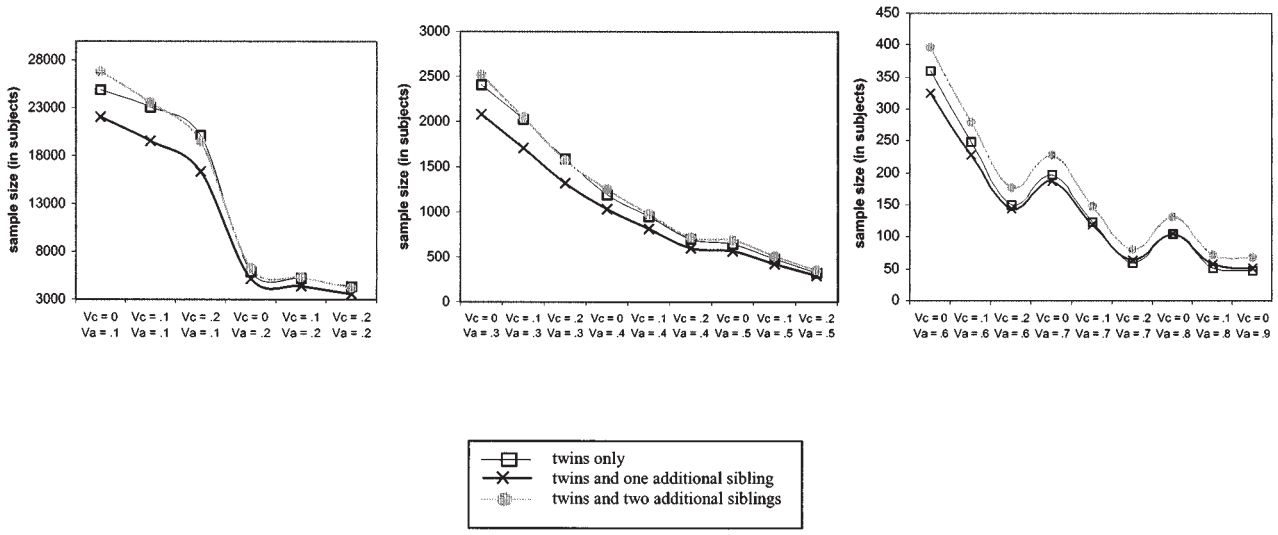


Fig. 2 a,b,c. Required sample size to detect sources of variance due to additive genetic effects in ACE models for three different family designs with a power of 80%. Design 1 = MZ and DZ twins only, Design 2 = MZ and DZ twins and one additional sibling, Design 3 = MZ and DZ twins and two additional siblings.

cerns low values of V_a (10% – 20%), Figure 2b concerns intermediate values of V_a (30%–50%), and Figure 2c concerns high values of V_a (60%–90%) accounting for the total phenotypic variance. All values of V_a are reported three times, i.e. in the context of values of V_c of 0%, 10%, and 20%.

As can be seen in Figure 2a, 2b, and 2c, for various values of V_a and V_c, design 2 (families consisting of MZ and DZ twins and one non-twin sibling) is the most optimal design to detect sources of variation due to A, i.e. with design 2 fewer subjects are required to achieve a power of 80% (see appendix II). The number of subjects needed to detect a fixed value of V_a is on average 9.3% more in the classical twin design (design 1) compared with a design with twins and one additional sibling. This can result in 2849 fewer subjects that are needed with design 2 to detect an additive genetic influence of 10% compared with the classical twin design.

Including families with twins and two additional sibs, is *less* powerful than including families with twins and one additional sibling, and also less powerful than including families with twins only for the detection of V_a; adding two siblings at the cost of the total number of MZ twins is disadvantageous, but adding one sibling is ideal.

Results for detecting common environmental influences are given in Figures 3a, 3b, and 3c, for low, moderate, and high values of V_c respectively (see also Appendix III).

Under various values of V_c and V_a, the power to detect sources of variation due to C rises substantially when one sibling is added to the classical twin design; on average 50.4% fewer subjects are needed as compared to the classical twin design (design 1). Adding two siblings decreases sample size even more, but not as dramatically as the decrease from no additional siblings to one additional sibling.

Many empirical studies suggest models in which sources of variation due to C are of less importance than sources of variation due to A (Plomin, DeFries, & McClearn, 1990). Therefore, we also calculated the sample size required to detect small values of V_c in the context of higher values of V_a. Figure 4 depicts the number of subjects needed to detect values of V_c of 10% and 20% in the context of values of V_a of 20% , 30%, 40% or 50% (Appendix IV).

As expected, sample size required to detect V_c with a power of 80% decreases as a result of higher values of V_c and higher values of V_a. Comparing the sample size required to detect sources of variation due to A (Figure 2b) with the sample size required to detect sources of variation due to C, shows that in the realistic situation where $V_a > V_c$ sources of variation due to C are very difficult to detect. Even if the sample size is large enough to detect sources of variation due to A, the small value of V_c may still go undetected. If for example the true model is an ACE-model with $V_a = 50\%$, $V_c = 20\%$, and $V_e = 30\%$, and the total sample size 328

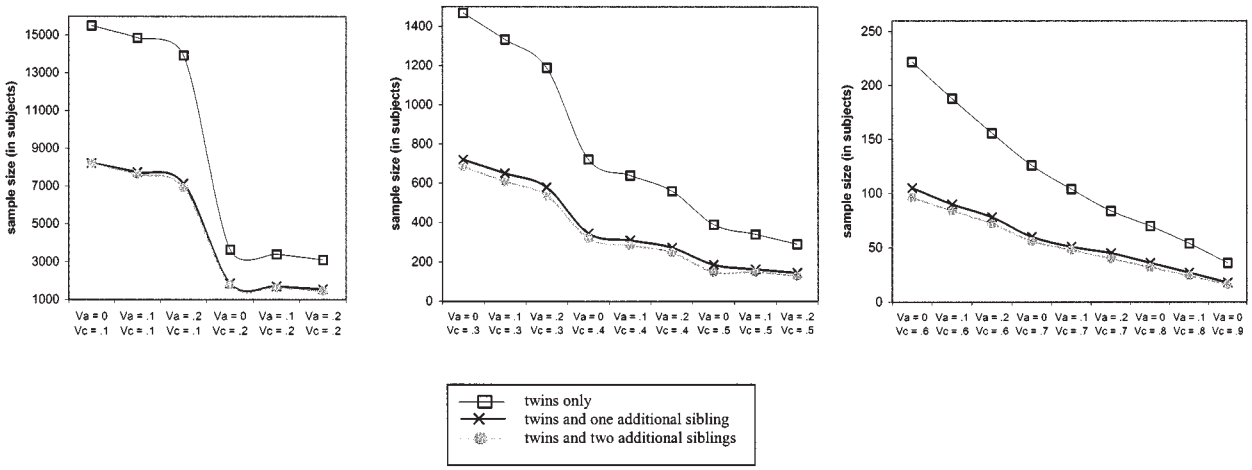


Fig. 3 a,b,c. Required sample size to detect sources of variance due to common environmental influences in ACE models for three different family designs with a power of 80%.

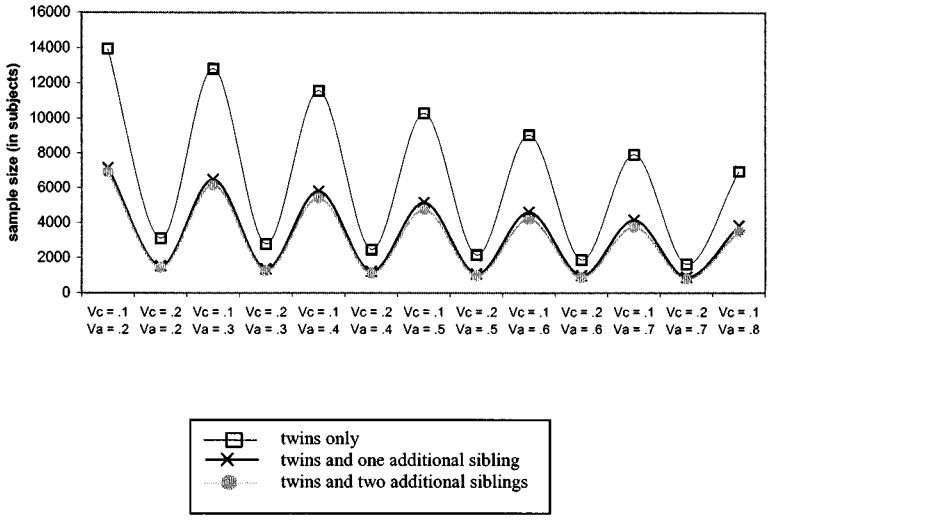


Fig. 4. Required sample size to detect sources of variance due to common environmental influences in ACE models where $V_a > V_c$, for three different family designs with a power of 80%.

(just enough for design 1 to detect V_a of 50%, with power of 80%), V_c will not be detected and the AE-model will be proposed as the most parsimonious model. This results in a biased estimate of V_a (in this case V_a is estimated to be 70%).

Adding siblings to the classical twin design decreases the sample size required to detect both V_a and V_c and has the largest effect on the sample size required to detect V_c (i.e. 50.4% fewer subjects needed for V_c, 9.3% fewer subjects needed for V_a). Therefore, the bias towards overestimating values of V_a as a result of not detecting V_c in situations where $V_a > V_c$, is less likely to be present in designs where siblings are added to the classical twin design.

ADE-Models

We also fitted full univariate models with sources of variation due to additive genetic (A), dominance (D) and unique environmental influences (E). Since a DE-model is unrealistic we report the sample size required to detect sources of variation due to A *and* D (2 df test) and to detect sources of variation due to D (1 df test) with a power of 80%. Results for detecting V_a and V_d, or V_d are given in Figures 5a and 5b (and appendix V).

Under various values of V_a and V_d, with fixed ratio of V_a to V_d is 2 to 1, adding one sibling to a twin family decreases the sample size required to detect V_d. Adding two siblings decreases sample size even more but less than the decrease due to adding one sibling. Absolute effects are slightly higher with increasing values of V_a and V_d. Figure 5a also emphasizes the very large sample size that is required to detect dominant genetic influences. Even the largest possible value of V_d under complete dominance with the most optimal design will go undetected if the sample is smaller than 1776 subjects.

Sample sizes required to detect both V_a and V_d simultaneously are considerably smaller as compared to sample sizes required to detect V_d. In contrast, however, adding siblings does not decrease sample size needed to detect V_a and V_d simultaneously. In fact, a design with one or two siblings requires somewhat more subjects to detect V_a and V_d with a power of 80%, as can be seen in Figure 5b. It should be noted however that the number of subjects needed to detect V_a and V_d at the same time is considerably less than the number of subjects needed to detect V_d only. This implies that if the sample size is large enough to detect V_d it will also be sufficient to detect V_a and V_d.

In conclusion, to optimize the power to detect V_d, a design with additional siblings, as compared to a design with twins only, is preferred.

Bivariate Models

ACE-Models

To detect sources of variance due to additive genetic influence (A), we calculated both the sample size required to detect all sources of V_a (df = 3; paths a_{11}, a_{21}, and a_{22} in Figure 1) and the required sample size to detect the common genetic pathway (df = 1; path a_{21}). We considered the test for the detection of the common pathway to be a test for the presence of a genetic correlation (rA). The following situations to detect sources of variance due to A were considered: a) The genetic correlation (rA) is 'moderate' and equal to the common environmental correlation (rC) and to the unique environmental correlation (rE). Variances due to A, C and E (uniform for both traits) are 40%, 10%, and 50% respectively of the phenotypic variance. b) rC is absent, rA is high (.80), and rE is small (.36), variances due to A, C and E are 40%, 10%, and 50% respectively.

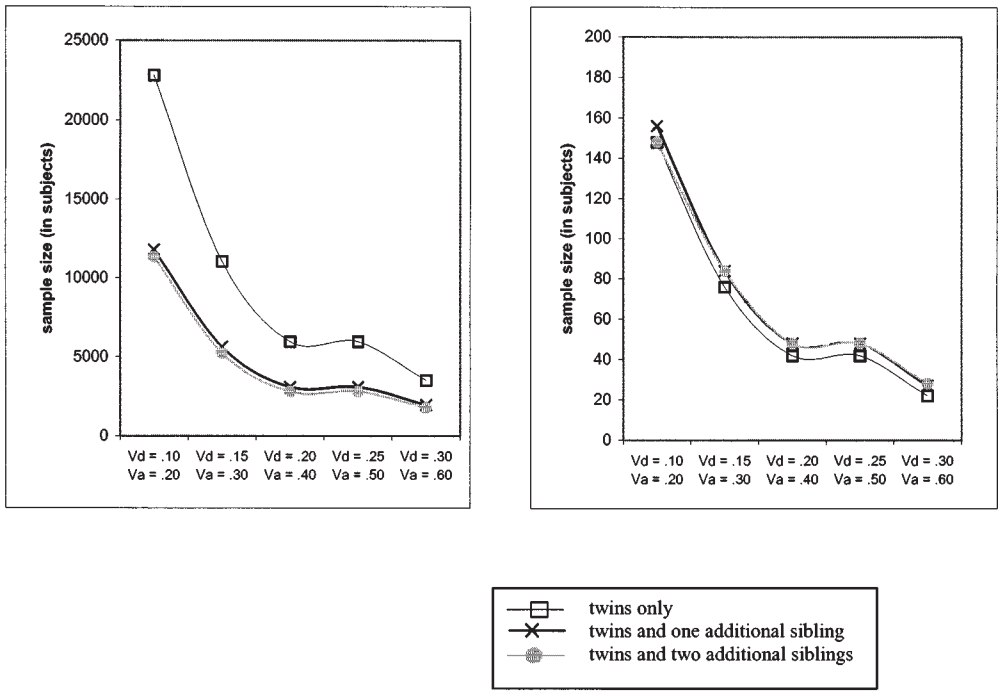


Fig. 5 a,b. Required sample size to detect sources of variance due to dominant genetic influences (a) and total genetic (dominant & additive influences)(b) influences in ADE models, for three different family designs with a power of 80%.

c) Variances due to C are absent. rA is .60, rE is .27, variances due to A and E are 70% and 30% respectively. As mentioned before, all parameters were estimated, as opposed to constraining these parameters, which were zero in the full model. It should also be noted that considering the tests for total V_a, total V_c, and total V_d to be 3 df-tests is a conservative approach, as it could be argued these are actually 2 df-tests, or tests with df's somewhere between 2 and 3. Testing, for example, whether either or both univariate genetic variances equal zero, implies that the genetic covariance is zero. If variances due to additive genetic influences for both traits equal zero, a correlation between these sources of variance is not possible. In other words, if the sample size required to detect each of the univariate variances due to additive genetic influences is insufficient, a correlation due to additive genetic influences can also not be detected. Therefore, considering the test for the power to detect 'total V_a' (i.e. both univariate variances due to additive genetic influences and the correlation due to additive genetic influences in the bivariate case) a 3 df test will provide an overestimation of the sample size needed for a power of 80%. Results of situation a, b, and c for the three different kinships, are given in Table I.

As can be seen in Table I the same pattern of results is found in the bivariate case as in the univariate case; a design with one additional sibling is optimal for the detection of V_a in ACE-models. In addition, significantly fewer subjects are needed in the bivariate case as compared to the univariate case. Depending on whether the phenotypic correlation is due to rA, rC, or rE, the sample size required to detect V_a may decrease and is lowest in cases where there is no influence of common environmental sources (i.e. statistical power is highest in these cases). However, when there are univariate common environmental influences but no common environmental correlation, the sample size required to detect variance due to additive genetic influences increases. Comparing situations a, b, and c leads to the conclusion that the power to detect sources of variance and covariance due to A (df 3) is highest (and the required sample size is smallest) when there is no univariate common environmental source of variation. However, if there are common environmental sources of variation, sources of variance due to A are easier to detect when there is also a correlation between these two univariate common environmental sources of variation, and again a design with one additional sibling is optimal.

To detect sources of common environmental sources of variation, we calculated both the power to detect all sources of variation due to C (df = 3) and the power to detect the common pathway (df = 1), which is a test to detect the environmental correlation (rC). We considered situations analogous to the situations in which power was calculated to detect sources of variation due to A; a) The common environmental correlation is 'moderate' and equal to the genetic correlation and to the unique environmental correlation, i.e. rC = rA = rE = .50. Uniform univariate variances due to A, C and E are 10%, 40%, and 50% respectively. b) rA is absent. rC is high (.80), and rE is small (.36), variances due to A, C and E are 10%, 40%, and 50% respectively, c) Variances due to A and rA are absent. rC is .60, rE is .27, variances due to C and E are 70% and 30% respectively. Again, for all situations the phenotypic correlation was .50. Results are given in Table II.

Although the results in the bivariate case resemble those in the univariate case (i.e. a design with two

Table I. Total samplesize (in number of subjects) needed to detect additive genetic influences in full bivariate ACE models under three different sibship sizes with power $(1 - \beta) = .80$ and $\alpha = .05$

	V_a = 40% rA = .50 V_c = 10% rC = .50 V_e = 50% rE = .50		V_a = 40% rA = .80 V_c = 10% rC = .00 V_e = 50% rE = .36		V_a = 70% rA = .60 V_c = 00% rC = .00 V_e = 30% rE = .27	
	all V_a (df = 3)	rA (df = 1)	all V_a (df = 3)	rA (df = 1)	all V_a (df = 3)	rA (df = 1)
design 1	660	2392	782	884	156	270
design 2	564	1917	678	735	147	237
design 3	680	2260	820	876	180	284

Note: MZ/DZ ratio = 1/1; design 1 = twins only, design 2 = twins and one additional sibling, design 3 = twins and two additional siblings. 'All V_a' refers to both univariate variances and the genetic correlation.
In order to calculate the total number of families needed, all cells concerning design 1 need to be divided by 2, all cells concerning design 2 need to be divided by 3, and all cells concerning design 3 need to be divided by 4.

Table II. Total samplesize (in number of subjects) needed to detect common environmental influences in full bivariate ACE models under three different sibship sizes with power $(1 - \beta) = .80$ and $\alpha = .05$

	$V_a = 10\%$ rA = .50 $V_c = 40\%$ rC = .50 $V_e = 50\%$ rE = .50		$V_a = 10\%$ rA = .00 $V_c = 40\%$ rC = .80 $V_e = 50\%$ rE = .36		$V_a = 0\%$ rA = .00 $V_c = 70\%$ rC = .60 $V_e = 30\%$ rE = .27	
	all V_c (df = 3)	rC (df = 1)	all V_c (df = 3)	rC (df = 1)	all V_c (df = 3)	rC (df = 1)
design 1	444	1498	518	560	100	156
design 2	213	774	249	279	48	96
design 3	48	760	44	268	16	108

Note: see table 1 for definitions.

additional siblings is optimal for the detection of V_c), the difference between design 2 and design 3 (i.e. adding one or two siblings) in the bivariate case is more substantial. Whereas in the univariate case only a small additional effect was found, in the bivariate case 4 to 5 times less subjects are needed with two additional siblings as compared to one additional sibling.

ADE-Models

We calculated covariance matrices for two traits that were influenced by A, D, and E in the context of complete dominance. Sources of variance due to A and D accounted for 40% and 20% respectively of the total phenotypic variance. We assumed that the ratio V_a to V_d remained equal over the two traits. This implies that rA = rD (see appendix I). Three situations were considered: a) rA = rD = .80,; b) rA = rD =.50; c) rA = rD = .30. For all three situations the phenotypic correlation was fixed at .50 by attributing all remaining covariance to rE. We report the total number of individual subjects needed to detect sources of total V_a and V_d due to A and D (df = 6), rA & rD (df = 2), total D (df = 3), and rD (df = 1) for a power of 80%. Results are given in Table III.

Analogous to the univariate case a design with two additional siblings is optimal for the detection of V_d and a design with twins only is optimal for the detection of V_a and V_d simultaneously. Comparison with the univariate results shows that in a design with twins only, fewer subjects are needed to detect sources of variance due to D as a result from bivariate testing. This effect, however, is stronger when a design consisting of twins and two additional siblings is used, suggesting that in addition to the decrease in sample size as a result from bivariate testing, adding siblings will decrease the sample size required to detect sources of variance due to D even further.

Designs Where Only Sibs of mz Twins are Included

In the previous analyses all families were of the same structure; consisting of MZ and DZ twins only, or with one or two additional siblings. For several reasons this may not always be realistic. For illustrative purposes, we included two other designs in which one (design 4) or two siblings (design 5) were added to MZ twin families, but not to DZ families. Analyses were run for a few 'standard' situations of the ACE-models and ADE-models for univariate testing only. Results for ACE and ADE models are given in Table IV.

Comparison of the results of designs 4 and 5 and the results of designs 2 and 3 shows that in ACE-models a design consisting of MZ twins and one additional sibling and DZ twins only (design 4) is optimal for the detection of V_a, and performs even better than design 2. For the detection of V_c in ACE-models design 3 and 5 are both optimal.

In the context of ADE-models, design 3 (MZ/DZ twins with two additional siblings), requires the smallest sample size and is more optimal than design 4 or 5 for the detection of sources of variation due to dominance.

CONCLUSION

We demonstrated that with a fixed power of 80%, a probablity level of 5% and under varying levels of heritability and common environmental influences, adding one sibling to the classical twin design significantly decreases the number of subjects that are needed to detect each of these sources of variation. Adding two siblings to a twin pair yields an additional decrease of sample size to detect sources of variation due to the common environment but is not optimal for the detection of additive genetic influences. If the trait is influenced by additive and non-additive genetic factors, adding one sibling to the classical twin design decreases the sample size needed to detect sources of variation

Table III. Total sample size (in subjects) needed to detect additive genetic influences and dominance in bivariate ADE models under under three different sibship sizes with power $(1 - \beta) = .80$ and $\alpha = .05$

	$V_a = 40\%$ rA = .80 $V_d = 20\%$ rD = .80 $V_e = 40\%$ rE = .05				$V_a = 40\%$ rA = .50 $V_d = 20\%$ rD = .50 $V_e = 40\%$ rE = .50				$V_a = 40\%$ rA = .30 $V_d = 20\%$ rD = .30 $V_e = 40\%$ rE = .80			
	all V_a + all V_d (df = 6)	rA&rD (df = 2)	all V_d (df = 3)	rD (df = 1)	all V_a + all V_d (df = 6)	rA&rD (df = 2)	all V_d (df = 3)	rD (df = 1)	all V_a + all V_d (df = 6)	rA&rD (df = 2)	all V_d (df = 3)	rD (df = 1)
design 1	62	64	8672	10042	54	184	7662	28054	32	548	4702	82356
design 2	69	78	4464	5121	60	228	3909	14073	39	681	2490	41544
design 3	72	88	4076	4660	60	264	3636	12872	40	784	2272	38436

Note: see table 1 for definitions.

Table IV. Total sample size (in number of subjects) needed to detect additive genetic, dominance and common environmental influences in univariate ACE-models and ADE-models for designs with including MZ and DZ twins and siblings added to MZ families only, a power of $(1 - \beta) = .80$, and significance level $\alpha = .05$

	$V_a = 40\%$ Vc = 10%	Vc = 40% $V_a = 10\%$	$V_a = 40\%$ $V_d = 20\%$	$V_a = 40\%$ $V_d = 20\%$
effect detected→	V_a	V_c	$V_a + V_d$	V_d
design 4	705	338	83	6313
design 5	744	285	84	5313

Note: MZ/DZ ratio = 1/1: design 4 = Mz twins and one additional sibling, DZ twins only, design 5 = MZ twins and two additional siblings and DZ twins onl.

due to dominance. Adding two siblings decreases the number of required subjects somewhat more but the decrease is relatively small (compared to the decrease due to adding one sibling). These effects are more pronounced in the bivariate case than in the univariate case. An additional benefit of adding siblings is that these designs, as compared to the classical twin design, are less likely to result in an overestimation of additive genetic influences as a result of not detecting small sources of common environmental influences.

We modeled the sibling covariances under the assumption that age differences in heritability are not important. A more complex model would take into account age differences between non-twin siblings. It is known that for some measures heritability increases with age as a result of amplification of genetic effects across ages (e.g. intelligence; Boomsma, 1993), whereas for other measures heritability estimates may decrease with age (e.g. problem behaviour; Van der Valk *et al.*, 1998). Assuming that the same genes operate across the age span, adding siblings who are older than the twins will increase power when heritability increase with age, and will decrease power when heritability estimates decrease with age. Similarly, adding parents will increase power to detect genetic factors if heritability increases with age.

Schork (1993) noted the dramatic improvement in statistical power resulting from the use of larger sibships for the detection of QTL effects. In addition, Dolan, Boomsma and Neale (1999) demonstrated the value of adding non-twin siblings to two-sibling- (or DZ twin-) families for the detection of codominant QTL effects. Our aim was to determine whether the use of an extended twin design, as needed for the detection of QTL-effects, would also be useful for the detection of overall

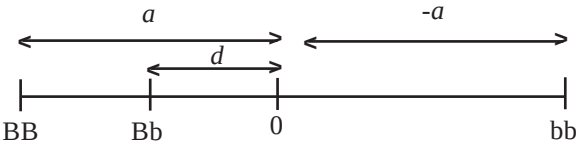
sources of variance (i.e. A, C, and D). Our calculations showed that without the need to increase total sample size, adding one sibling to the classical twin design improves the statistical power by a large extent to detect sources of variation due to common environmental influences, additive genetic influences and dominance. Adding siblings and using a bivariate phenotype results in gain of statistical power which can not only be ascribed to bivariate testing but also to the use of an extended twin design.

In conclusion, adding at least one sibling to the classical twin design, as opposed to a design with twins only, will provide a significant gain in statistical power to detects sources of variation due to A, C, and D. An attractive side-effect of a design with additional siblings is that it is also beneficial for the detection of QTL-effects.

APPENDIX I

Consider a biallelic trait with alleles B and b. Let *a* be the effect of genotype BB on the phenotypic mean, *–a* the effect of bb, and *d* the effect of Bb on the phenotypic mean. Assuming equal allele frequencies of B and b, the mean genotypic effect on the phenotypic mean is *1/2 d*. The total genetic variance ($\sigma^2 g$) equals $1/2\ a^2 + 1/4\ d^2$, $= V^a + V^d$

For complete dominance d = a. Substituting d for a in the formulae for the genetic variances, gives: $V_a = 1/2\ a^2$ and $V^d\ 1/4\ a^2$, thus $V^a = 2\ V^d$



Now consider a *bivariate* model with latent variances scaled to unity, (see figure 1) and

- uniform genetic influences over traits: $V_{a\,1} = V_{a\,2}$ and $V_{d\,1} = V_{d\,2}$
- assumption of uniform d to a ratio over traits $(a_{11})^2/(d_{11})^2 = (a_{21})^2/(d_{21})^2 = (a_{22})^2/(d_{22})^2$
- $rA = a_{11} * a_{21}/\sqrt{\{(a_{11})^2 * [(a_{21})^2 + [(a_{22})^2]\}}$ which simplifies to $rA = a_{21}/a_{11}$
- $rD = d_{11} * d_{21}/\sqrt{\{(d_{11})^2 * [(d_{21})^2 + [(d_{22})^2]\}}$ which simplifies to $rD = d_{21}/d_{11}$

This implies that the additive genetic correlation equals the dominant genetic correlation.

APPENDIX II

Sample size (in subjects) needed to detect additive genetic influences in full univariate ACE models under varying levels of variation due to common environmental sources for three different sibshipsizes.
MZ/DZ ratio = 1/1, significance level $\alpha = .05$, power $(1 - \beta) = .80$, design 1 = twins only, design 2 = twins and one additional sibling, design 3 = twins and two additional siblings.
In order to calculate the total number of families needed, all cells from design 1 need to be divided by 2, all cells from design 2 need to be divided by 3, and all cells from design 3 need to be divided by 4.

	$V_a = 10\%$			$V_a = 20\%$			$V_a = 30\%$			$V_a = 40\%$			$V_a = 50\%$			$V_a = 60\%$			$V_a = 70\%$			$V_a = 80\%$		$V_a = 90\%$
$V_c \rightarrow$	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	0%
design 1	24896	23084	20110	5908	5230	4332	2406	2026	1588	1192	950	700	644	482	328	360	248	150	198	124	60	104	52	48
design 2	22047	19557	16365	5151	4395	3540	2079	1707	1320	1032	813	600	567	426	294	324	228	144	186	120	63	105	57	51
design 3	26836	23560	19460	6256	5280	4208	2520	2048	1572	1252	976	716	688	512	356	396	280	176	228	148	80	132	72	68

APPENDIX III

Sample size (in subjects) needed to detect common environmental influences in full univariate ACE models under varying levels of variation due to additive genetic sources for three different sibshipsizes.

See Appendix II for definitions

	V_c = 10%			V_c = 20%			V_c = 30%			V_c = 40%			V_c = 50%			V_c = 60%			V_c = 70%			V_c = 80%		V_c = 90%
$V_a \rightarrow$	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	20%	0%	10%	0%
design 1	15504	14860	13934	3646	3398	3104	1468	1334	1190	722	640	560	390	340	290	222	188	156	126	104	84	70	54	36
design 2	8220	7746	7140	1860	1704	1542	720	651	579	345	309	273	186	162	144	105	90	78	60	51	45	36	27	18
design 3	8220	7628	6908	1808	1632	1448	684	608	536	320	284	248	148	148	128	96	84	72	56	48	40	32	24	16

APPENDIX IV

Sample size (in subjects) needed to detect common environmental influences in full univariate ACE models under varying levels of variation due to additive genetic sources in the realistic situation that sources of variation due to A are larger than sources of variation due to C for three different sibshipsizes.

See Appendix II for definitions

	V_a = 20%		V_a = 30%		V_a = 40%		V_a = 50%		V_a = 60%		V_a = 70%		V_a = 80%
$V_c \rightarrow$	10%	20%	10%	20%	10%	20%	10%	20%	10%	20%	10%	20%	10%
design 1	13940	3104	12806	2786	11558	2466	10280	2158	9042	1876	7912	1628	6934
design 2	7143	1542	6471	1377	5790	1224	5151	1089	4590	978	4137	894	3822
design 3	6912	1448	6148	1276	5408	1120	4736	988	4172	884	3748	816	3496

APPENDIX V

Samplesize (in subjects) needed to detect additive genetic and dominance influences in ADE-models. *See Appendix II for definitions*

	V_a = 20% V_d = 10%		V_a = 30% V_d = 15%		V_a = 40% V_d = 20%		V_a = 50% V_d = 25%		V_a = 60% V_d = 30%	
	V_a & V_d	V_d	V_a & V_d	V_d	V_a & V_d	V_d	V_a & V_d	V_d	V_a & V_d	V_d
design 1	148	22808	76	11036	42	5958	42	5958	22	3518
design 2	156	11790	84	5631	48	3081	48	3081	27	1950
design 3	148	11328	84	5236	48	2784	48	2784	28	1776

REFERENCES

Boomsma, D. I. Current status and future prospects in twin studies of the development of cognitive abilities: Infancy to old age. In: Bouchard, Thomas J. Jr. (Ed), Propping, Peter (Ed), *et al.* (1993). *Twins as a tool of behavioral genetics. Life sciences research report,* **53**. (pp. 67–82). Chichester, England UK: John Wiley & Sons.

Boomsma, D. I. & Dolan, C. V. (1998). A comparison of power to detect a QTL in sib-pair data using multivariate phenotypes, mean phenotypes, and factor scores. *Behavior Genetics,* **28**:329–340.

Cohen, J. (1992) A power primer. *Psychological Bulletin,* **112**:155–159.

Dolan, C. V., Boomsma, D. I., & Neale, M. C. (1999). A note on the power provided by sibships of size 2, 3, and 4 in genetic covariance modeling of a codominant QTL. *In press*

Fulker, D. W., Cherny, S. S., Cardon, L. R. (1995). Multipoint interval mapping of Quantitative Trait Loci, using sib pairs. *Am J Hum Genet,* **56**(5):1224–1233

Fulker, D. W., Cherny, S. S., Sham, P. C., *et al.*(1999). Combined linkage and association. Sib-pair analysis for quantitative traits. *Am J Hum Genet,* **64**(1):259–267

Hewitt, J. K., & Heath, A. C. (1988). A note on computing the chi-square noncentrality parameter for power analyses. *Behavior Genetics,* **18**:105–108.

Martin, N. G., Eaves, L. J., Kearsey, M. J., and Davies, P. (1978). The power of the classical twin study. *Heredity,* **40**:97–116.

Nance, W. E., & Neale, M. C. (1989). Partitioned twin analyses: a power study. *Behavior Genetics,* **19**:143–150.

Neale, M. C. (1997). *Mx: Statistical modeling.* 3rd edition Box 980126 MCV, Richmond VA 23298.

Neale, M. C., & Cardon L. R.(1992) *Methodology for Genetic Studies of Twins and Families.* NATO Asi Series. Series D,Behavioural and Social Sciences, Vol 67.

Pickens, R. W., Svikis, D. S., McGue, M., Lykken, D. T. *et al.* (1991). Heterogeneity in the inheritance of alcoholism: A study of male and female twins. *Archives of General Psychiatry,* **48**:19–28.

Plomin, R., DeFries, J. C., & McClearn, G. E. (1990) *Behavioral Genetics. A primer.* New York: Freeman and company.

Schmitz, S. Cherny, S. S., & Fulker, D. W. (1998). Increase in power through multivariate analyses. *Behavior Genetics,* **28**:357–364.

Schork, N. J., (1993). Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *American Journal of Human Genetics,* **53**:1306–1319.

Svikis, D. S., Velez, M. L., Pickens, R. W. (1994). Genetic aspects of alcohol use and alcoholism in women. *Alcohol Health & Research World,* **18**:192–196.

Tanaka, J. S. (1987). How big is big enough?: sample size and goodness of fit in structural equation models with laten variables. *Child Development,* **58**:134–146.

van der Valk, J. C., Verhulst, F. C., Neale, M. C., Boomsma, D. I. (1998). Longitudinal genetic analysis of problem behaviors in biologically related and unrelated adoptees. *Behavior Genetics,* **28**:365–380.

Edited by Norman D. Henderson